

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-117046

(43)Date of publication of application : 19.04.2002

(51)Int.Cl.

G06F 17/30

(21)Application number : 2000-306537

(71)Applicant : JUST SYST CORP

(22)Date of filing : 05.10.2000

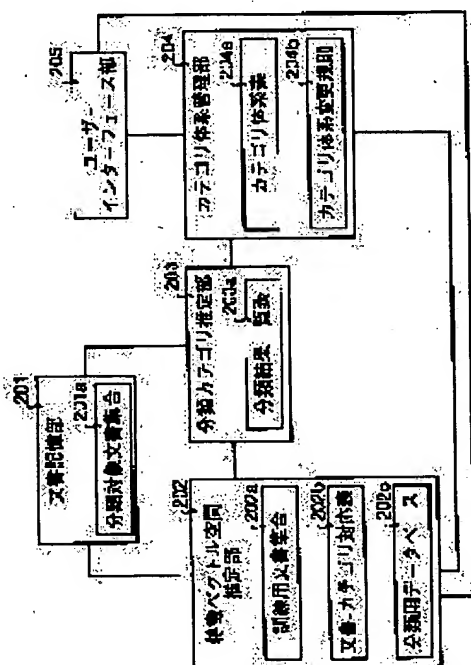
(72)Inventor : TSUJII MASAKI

(54) DEVICE AND METHOD FOR CLASSIFYING DOCUMENT, AND COMPUTER-READABLE RECORDING MEDIUM RECORDING PROGRAM FOR MAKING COMPUTER IMPLEMENT THE SAME METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To prevent classification precision from becoming worse as a group of documents to be classified changes in quality by making a document classifying device which classifies a computerized document into one of previously set categories always maintain a classification system matching with characteristics of the group of the documents to be classified.

SOLUTION: This device is equipped with a feature vector space estimation part 202 which estimates feature vector spaces of respective categories (including an unclassifiable category) into which documents are classified, a classification category estimation part 203 which estimates classification destinations of the respective documents by comparing feature vectors of the documents with the feature vector spaces of the respective categories, a category system management part 204 which decides whether the frequency of estimation that an unclassifiable category is estimated as a classification destination exceeds a certain threshold, and a user interface part 205 which recommends an operator to add a new category when the threshold is exceeded. The categories can be divided, deleted, or merged.



BEST AVAILABLE COPY

## LEGAL STATUS

[Date of request for examination] 05.10.2000

[Date of sending the examiner's decision of rejection] 08.06.2004

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-117046

(P2002-117046A)

(43) 公開日 平成14年4月19日 (2002.4.19)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30	2 1 0	G 0 6 F 17/30	2 1 0 D 5 B 0 7 5
	1 7 0		1 7 0 A
	3 4 0		3 4 0 B
	3 5 0		3 5 0 C

審査請求 有 請求項の数 7 O L (全 20 頁)

(21) 出願番号 特願2000-306537(P2000-306537)

(22) 出願日 平成12年10月5日 (2000.10.5)

(71) 出願人 390024350

株式会社ジャストシステム

徳島県徳島市沖浜東3-46

(72) 発明者 辻井 雅樹

徳島市沖浜東3丁目46番地 株式会社ジャストシステム内

(74) 代理人 100104190

弁理士 酒井 昭徳

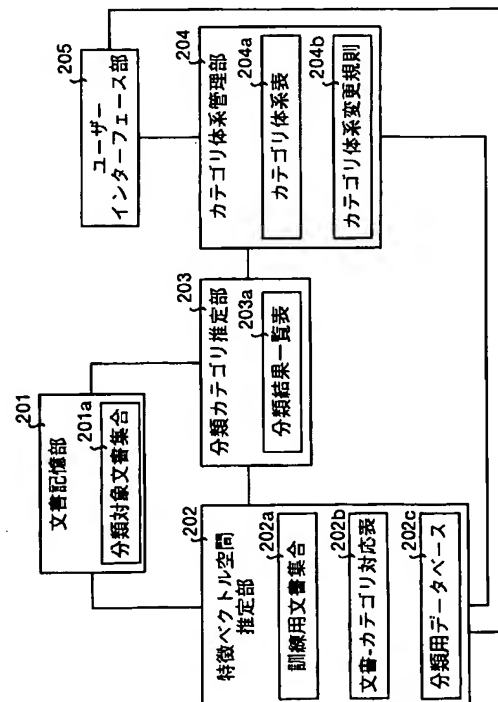
Fターム(参考) 5B075 NK46 NR12 PP03 PQ46 PR04  
UU06

(54) 【発明の名称】 文書分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体

## (57) 【要約】

【課題】 電子化された文書をあらかじめ設定されたカテゴリのいずれかに分類する文書分類装置において、分類対象文書群の特性に見合った分類体系を常に維持できるようにし、分類対象文書群の質的变化にともなう分類精度の低下を防止すること。

【解決手段】 文書が分類される各カテゴリ（分類不能カテゴリを含む）の特徴ベクトル空間を推定する特徴ベクトル空間推定部202と、各文書の特徴ベクトルと各カテゴリの特徴ベクトル空間とを比較して各文書の分類先を推定する分類カテゴリ推定部203と、分類不能カテゴリが分類先として推定される頻度が一定の閾値を上回ったかどうかを判定するカテゴリ体系管理部204と、閾値を上回った場合に新たなカテゴリの追加を操作者に対して推奨するユーザーインターフェース部205とを備える。カテゴリの分割・削除あるいは併合も可能である。



## 【特許請求の範囲】

【請求項1】 電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置において、

訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する第1の特徴ベクトル空間推定手段と、

すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定する第2の特徴ベクトル空間推定手段と、

前記第1の特徴ベクトル空間推定手段により推定された各カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定手段により推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定手段と、

前記分類カテゴリ推定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定するカテゴリ追加要否判定手段と、前記カテゴリ追加要否判定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するカテゴリ追加推奨手段と、を備えたことを特徴とする文書分類装置。

【請求項2】 電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置において、

訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する特徴ベクトル空間推定手段と、

前記特徴ベクトル空間推定手段により推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定手段と、

前記分類カテゴリ推定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定するカテゴリ削除・併合要否判定手段と、前記カテゴリ削除・併合要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するカテゴリ削除・併合推奨手段と、

を備えたことを特徴とする文書分類装置。

【請求項3】 さらに、前記分類カテゴリ推定手段により分類先のカテゴリと推定される頻度が一定の閾値を上

回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定手段と、

前記カテゴリ分割要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨手段と、

を備えたことを特徴とする前記請求項1または2に記載の文書分類装置。

【請求項4】 電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、

訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する第1の特徴ベクトル空間推定工程と、

すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定する第2の特徴ベクトル空間推定工程と、

前記第1の特徴ベクトル空間推定工程で推定された各カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定工程で推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定工程と、

前記分類カテゴリ推定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定するカテゴリ追加要否判定工程と、

前記カテゴリ追加要否判定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するカテゴリ追加推奨工程と、を含んだことを特徴とする文書分類方法。

【請求項5】 電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、

訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する特徴ベクトル空間推定工程と、

前記特徴ベクトル空間推定工程で推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定工程と、

前記分類カテゴリ推定工程で分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定するカテゴリ削除・併合要否判定工程と、

前記カテゴリ削除・併合要否判定工程で分類先のカテゴリ

りと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するカテゴリ削除・併合推奨工程と、

を含んだことを特徴とする文書分類方法。

【請求項6】 さらに、前記分類カテゴリ推定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定工程と、

前記カテゴリ分割要否判定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨工程と、

を含んだことを特徴とする前記請求項4または5に記載の文書分類方法。

【請求項7】 前記請求項4～6のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【0002】

【従来の技術】コーパスと呼ばれる文書群をもちいて、与えられた分類対象文書（群）をあらかじめ設定されたいずれかのカテゴリに自動分類する文書分類装置が従来から知られている。この種の装置では、あらかじめ多種多様な内容を有する大量の文書（コーパス）について、人手によりそれぞれの分類先のカテゴリを決定しておき、各カテゴリに分類される文書の特徴から逆に各カテゴリの特徴を算出しておく。そして、分類対象文書が与えられたときは、当該文書の特徴と各カテゴリの特徴とを順次比較して、もっとも類似度の高いカテゴリへ当該文書を分類する。

【0003】ただ、分類される文書の特性によっては、あらかじめ設定された分類体系が不適切となる場合もあるので、分類対象文書の傾向にあわせて、自動的に分類体系を変更してゆく文書分類装置も提案されている。このような従来技術としては、たとえば特開平07-049875や特開2000-011004を挙げることができる。

【0004】特開平07-049875における文書収集サーバシステムは、自動的に複数の情報源に接続して新文書を取得し、適合度計算によって、あらかじめユーザーが記述した検索条件との適合度を調べる。そして、検索条件間の関係から分類体系を構成し、適合した文書

を分類してフォルダに格納する。さらに、各フォルダへの情報の集まり具合を監視し、自動的にフォルダの細分化、統合、構造の変更をおこなう。

【0005】また、特開2000-011004における情報自動分類装置は、あるカテゴリに誤って分類された文書がある場合に、当該カテゴリに付随する仮カテゴリを設けてそこに上記文書を分類するようにしている。そして、以後の文書は正規のカテゴリまたは仮カテゴリのうちより類似度の高い方に分類される。すなわち、既存のカテゴリの範疇を超える文書が出現した場合には、正規のカテゴリに隣接してそれに準ずる仮カテゴリが自動的に生成されることになる。この仮カテゴリを正規のカテゴリに格上げすることもできる。

【0006】

【発明が解決しようとする課題】しかしながら、上記従来技術のうち特開平07-049875においては、既存のカテゴリを分割（細分化）してゆくことができるのみなので、まったく新しい分野の文書が出現した場合には当該文書にふさわしいカテゴリを生成することはできない。また、文書がほとんどあるいはまったく分類されていない、不必要なカテゴリを削除することもできない。また、特開2000-011004によっても、既存のカテゴリとの類似性の低いカテゴリを生成することはできず、またカテゴリの分割や削除もおこなうことができない。

【0007】すなわち、上記従来技術は分類対象文書群の質的变化に弱く、既存のカテゴリに分類できない新規な文書が現れてくると、分類体系が分類対象文書群の特性にそぐわないものである結果、適切な分類ができず分類精度が低下してしまうという問題点があった。

【0008】この発明は、上述した従来技術による問題点を解消するため、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【0009】

【課題を解決するための手段】上述した課題を解決し、目的を達成するため、請求項1に記載の発明にかかる文書分類装置は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置において、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する第1の特徴ベクトル空間推定手段と、すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定する第2の特徴ベクトル空間推定手段と、前記第1の特徴ベクトル空間推定手段により推定された

各カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定手段により推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定手段と、前記分類カテゴリ推定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定するカテゴリ追加要  
10 判定手段と、前記カテゴリ追加要否判定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するカテゴリ追加推奨手段と、を備えたことを特徴とする。

【0010】この請求項1に記載の発明によれば、分類不能カテゴリに分類される文書が多くなると、操作者に対して新たなカテゴリの追加が推奨される。

【0011】また、請求項2に記載の発明にかかる文書分類装置は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置  
20 において、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する特徴ベクトル空間推定手段と、前記特徴ベクトル空間推定手段により推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定手段と、前記分類カテゴリ推定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定するカテゴリ削除・併合  
30 要否判定手段と、前記カテゴリ削除・併合要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するカテゴリ削除・併合推奨手段と、を備えたことを特徴とする。

【0012】この請求項2に記載の発明によれば、あるカテゴリに分類される文書が少なくなると、操作者に対して当該カテゴリの削除または併合が推奨される。

【0013】また、請求項3に記載の発明にかかる文書分類装置は、前記請求項1または2に記載の発明において、さらに、前記分類カテゴリ推定手段により分類先の  
40 カテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定手段と、前記カテゴリ分割要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨手段と、を備えたことを特徴とする。

【0014】この請求項3に記載の発明によれば、ある

カテゴリに分類される文書が多くなると、操作者に対して当該カテゴリの分割が推奨される。

【0015】また、請求項4に記載の発明にかかる文書分類方法は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する第1の特徴ベクトル空間推定工程と、すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定する第2の特徴ベクトル空間推定工程と、前記第1の特徴ベクトル空間推定工程で推定された各  
50 カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定工程で推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定工程と、前記分類カテゴリ推定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定するカテゴリ追加要否判定工程と、前記カテゴリ追加要否判定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するカテゴリ追加推奨工程と、を含んだことを特徴とする。

【0016】この請求項4に記載の発明によれば、分類不能カテゴリに分類される文書が多くなると、操作者に対して新たなカテゴリの追加が推奨される。

【0017】また、請求項5に記載の発明にかかる文書分類方法は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定する特徴ベクトル空間推定工程と、前記特徴ベクトル空間推定工程で推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定する分類カテゴリ推定工程と、前記分類カテゴリ推定工程で分類先の  
60 カテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定するカテゴリ削除・併合要否判定工程と、前記カテゴリ削除・併合要否判定工程で分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するカテゴリ削除・併合推奨工程と、を含んだことを特徴とする。

【0018】この請求項5に記載の発明によれば、ある  
50 カテゴリに分類される文書が少なくなると、操作者に対

して当該カテゴリの削除または併合が推奨される。

【0019】また、請求項6に記載の発明にかかる文書分類方法は、前記請求項4または5に記載の発明において、さらに、前記分類カテゴリ推定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定工程と、前記カテゴリ分割要否判定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨工程と、を含んだことを特徴とする。

【0020】この請求項6に記載の発明によれば、あるカテゴリに分類される文書が多くなると、操作者に対して当該カテゴリの分割が推奨される。

【0021】また、請求項7に記載の発明にかかる記録媒体は、前記請求項4～6のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことで、当該プログラムをコンピュータで読み取ることが可能となり、これによって、請求項4～6のいずれか一つに記載された方法をコンピュータによって実施することが可能となる。

【0022】

【発明の実施の形態】以下に添付図面を参照して、この発明にかかる文書分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体の好適な実施の形態を詳細に説明する。

【0023】（実施の形態）まず、この発明の実施の形態による文書分類装置のハードウェア構成について説明する。図1は、この発明の実施の形態による文書分類装置のハードウェア構成を示す説明図である。同図において、101はシステム全体を制御するCPUを、102は基本入出力プログラムを記憶したROMを、103はCPU101のワークエリアとして使用されるRAMを、それぞれ示している。

【0024】また、104はCPU101の制御にしたがってHD（ハードディスク）105に対するデータのリード／ライトを制御するHDD（ハードディスクドライブ）を、105はHDD104の制御にしたがって書き込まれたデータを記憶するHDを、それぞれ示している。また、106はCPU101の制御にしたがってFD（フロッピー（登録商標）ディスク）107に対するデータのリード／ライトを制御するFDD（フロッピーディスクドライブ）を、107はFDD106の制御にしたがって書き込まれたデータを記憶する着脱自在のFDを、それぞれ示している。

【0025】また、108はカーソル、メニュー、ウィンドウ、あるいは文字や画像などの各種データを表示するディスプレイを、109は通信回線110を介してネットワークNETに接続され、そのネットワークNET

とCPU101とのインターフェースとして機能するネットワークボードを、それぞれ示している。また、111は文字、数値、各種指示などの入力のための複数のキーを備えたキーボードを、112は各種指示の選択や実行、処理対象の選択、カーソルの移動などをおこなうマウスを、それぞれ示している。

【0026】また、113は文字や画像を光学的に読み取るスキャナを、114はCPU101の制御にしたがって文字や画像を印刷するプリンタを、115は着脱可能な記録媒体であるCD-ROMを、116はCD-ROMドライブを、100は上記各部を接続するためのバスまたはケーブルを、それぞれ示している。

【0027】つぎに、この発明の実施の形態にかかる文書分類装置の機能的構成について説明する。図2は、この発明の実施の形態にかかる文書分類装置の構成を機能的に示す説明図である。

【0028】201は、文書記憶部であり、後述する分類カテゴリ推定部203による分類の対象となる文書群（分類対象文書集合）201aを保持している。それぞれの文書は、ネットワーク上の他の情報処理装置などから通信回線110を介して取得されたり、キーボード111によって入力されたり、紙媒体の文書からスキャナ113を介して取り込まれたり、あるいはCD-ROM115などの各種記録媒体から読み込まれたりしたものである。なお、それぞれの文書の取得・蓄積時に差異があってもよい（50個の文書のうち、10個は一週間前に取得・蓄積され、25個は昨日取得・蓄積され、15個は今日取得・蓄積された、など）。

【0029】202は、特徴ベクトル空間推定部であり、初期状態では訓練用文書集合202aと文書－カテゴリ対応表202bとを保持している。訓練用文書集合202aはいわゆるコーパスであって、様々な分野に属する文書をまんべんなく含む文書集合、たとえば一年分の新聞記事からなる文書集合である。また、文書－カテゴリ対応表202bは、訓練用文書集合202aの各文書に対応づけて、当該文書が分類されるカテゴリを記憶したテーブルである。このカテゴリの名称や上位・下位関係は、後述するカテゴリ体系管理部204のカテゴリ体系表204aで定義されている。

【0030】特徴ベクトル空間推定部202は、訓練用文書集合202aと文書－カテゴリ対応表202bとから、分類用データベース202cを作成する。図3は、分類用データベース202cの作成方法を具体的に説明するための説明図である。訓練用文書集合202aが図3（a）に示すD1～D10の10個の文書からなり、それらの文書が図3（b）の文書－カテゴリ対応表202bに示すように、A、BまたはCのいずれかのカテゴリに分類されていたものとする。

【0031】特徴ベクトル空間推定部202は、まず訓



練用文書集合202aの個々の文書について、当該文書の意味内容を近似的に表現するベクトル（特徴ベクトル）を作成する。特徴ベクトルの作成方法としては種々の従来技術が存在するが、ここでは説明の便宜上、文書集合内に出現する全単語（全語彙。ただし不要語などを除く）の各文書内での出現頻度（出現回数）を計数して、ベクトル内の各要素値（各ベクトルの長さ）とするもっとも単純な方法を採用する。

【0032】訓練用文書集合202a内に出現する単語がW1～W5の5種類であった場合、各文書について作成される特徴ベクトルは、当該文書内での単語W1の出現頻度、W2の出現頻度、W3の出現頻度、W4の出現頻度、およびW5の出現頻度を各ベクトル長さとする5次元のベクトルとなる。図3(c)に、各文書について作成される特徴ベクトルの一例を示す。たとえば、単語W1が0回、W2が2回、W3が1回、W4が0回、W5が3回出現する文書D1の特徴ベクトルは、 $V1 = (0, 2, 1, 0, 3)$ である。同様に、文書D5の特徴ベクトルは $V5 = (1, 5, 1, 0, 4)$ 、文書D8の特徴ベクトルは $V8 = (0, 2, 0, 0, 2)$ である。

【0033】訓練用文書集合202aの各文書についてその特徴ベクトルを作成すると、つぎに特徴ベクトル空間推定部202は、各カテゴリに分類される文書の特徴ベクトルを平均することで、各カテゴリの特徴ベクトル空間を算出する。たとえば、カテゴリAの特徴ベクトル空間VAは、カテゴリAに属する文書D1、D5およびD8の特徴ベクトルの平均を取って、 $VA = (0+1+0/3, 2+5+2/3, 1+1+0/3, 0+0+0/3, 3+4+2/3)$ 、すなわち $VA = (0.33, 3, 0.66, 0, 3)$ となる。そしてこのVA、また同様にして求めたVBおよびVCから、図3(d)に示す分類用データベース202cを作成する。

【0034】ここで、分類用データベース202cには、A、BおよびCのほかにカテゴリ「分類不能」が設けられている。この分類不能カテゴリの特徴ベクトル空間は、以下の(1)～(4)の手順にしたがって算出する。

【0035】(1)すべてのベクトルの長さが等しい特徴ベクトル空間VAの算出

まず、特徴ベクトル空間推定部202は、文書記憶部201に保持されている分類対象文書集合201aの各文書について、当該文書集合内に出現する全単語、あるいはあらかじめ選定されたいくつかの単語の、各文書内での出現頻度を要素値とする特徴ベクトルを作成する。ここでは、分類対象文書集合201aが図4(a)に示すD'1～D'5の5個の文書からなり、各文書について図4(b)に示すような特徴ベクトルが作成されたものとする。

【0036】そして、特徴ベクトル空間推定部202

は、訓練用文書集合202aおよび分類対象文書集合201aにおいて単語W1～W7が出現する全回数（ベクトル長さの合計といってもよい。ここでは $87+59=146$ ）を、全文書数（ここでは $10+5=15$ ）で除算して、一文書あたりの平均単語出現回数（ここでは $146/15=9.73$ ）を算出する。さらに、これを全語彙数（ベクトルの数といってもよい。ここでは5）で除算して、一文書あたりの各単語の平均出現回数（ここでは $9.73/5=1.95$ ）を算出する。そして、この値を単語W1～W7に対応する各ベクトルの長さとする特徴ベクトル空間 $V\alpha = (1.95, 1.95, 1.95, 1.95, 1.95, 1.95, 1.95)$ を作成する。

【0037】(2)訓練用文書集合202aの特徴ベクトル空間Vβの算出

つぎに、特徴ベクトル空間推定部202は、単語W1～W7のそれぞれが訓練用文書集合202a内の一文書あたりに平均何回出現しているかを算出して、訓練用文書集合202aの特徴ベクトル空間Vβを作成する。ここでは $V\beta = (12/10, 11/10, 13/10, 17/10, 34/10, 0/10, 0/10)$ 、すなわち $V\beta = (1.2, 1.1, 1.3, 1.7, 3.4, 0, 0)$ となる。

【0038】(3)分類対象文書集合201aの特徴ベクトル空間Vyの算出

さらに、特徴ベクトル空間推定部202は、単語W1～W7のそれぞれが分類対象文書集合201a内の一文書あたりに平均何回出現しているかを算出して、分類対象文書集合201aの特徴ベクトル空間Vyを作成する。ここでは $Vy = (7/5, 6/5, 8/5, 10/5, 18/5, 10/5, 0/5)$ 、すなわち $Vy = (1.4, 1.2, 1.6, 2, 3.6, 2, 0)$ となる。

【0039】(4)分類不能カテゴリの特徴ベクトル空間Vの算出

そして、特徴ベクトル空間推定部202は、上記で算出したVα、VβおよびVyの重み付き平均を取ること、分類不能カテゴリの特徴ベクトル空間Vを算出する。たとえば、Vαに与える重みが0.15、Vβに与える重みが0.45、Vyに与える重みが0.40であったとすると、 $V = ((1.95 \times 0.15) + (1.2 \times 0.45) + (1.4 \times 0.40), (1.95 \times 0.15) + (1.1 \times 0.45) + (1.2 \times 0.14), \dots)$ 、すなわち $V = (1.39, 1.27, 1.52, 1.86, 3.26, 1.09, 0.29)$ となる。そして、このVを分類不能カテゴリの特徴ベクトル空間として、分類用データベース202cに登録する。

【0040】このように、分類不能カテゴリの特徴ベクトル空間Vを「すべてのベクトルの長さが等しい特徴ベクトル空間Vα」「訓練用文書集合202aの特徴ベク

トル空間 $V\beta$ 」および「分類対象文書集合201aの特徴ベクトル空間 $V\gamma$ 」の加重平均によって推定することの意味は以下の通りである。

【0041】まず、すべてのベクトルの長さが等しい特徴ベクトル空間 $V\alpha$ は、未知のカテゴリの特徴ベクトル空間を仮定する役割を担っている。未知のカテゴリでは、各語彙がどのような分布を取るかが不明なので、すべての語彙の出現頻度が同じというベクトル空間をこの $V\alpha$ によって仮定する。

【0042】もっとも、この $V\alpha$ は自然な言語空間の表現とはなっていない。たとえば、英語の「the」などはカテゴリに関係なくどの文書でも出現頻度が高いため、分類不能カテゴリにおいても頻度が高くなっているのが自然である。そこで、訓練用文書集合202aの特徴ベクトル空間 $V\beta$ を考慮することにより、出現頻度がカテゴリに依存しない語（図3の例ではW5）のベクトル長さを改善することができる。

【0043】分類対象文書集合201aの特徴ベクトル空間 $V\gamma$ を考慮に入れるのも、基本的には上記と同様である。ただし、訓練用文書集合202aはその規模に限りがあるので（あらかじめ各文書の内容を検討して、人手によってカテゴリを付与しなければならないため）、その規模を補う意味で、 $V\beta$ に加えて $V\gamma$ を考慮する。また、訓練用文書を収集したときと語彙のパターンが変わった場合など、 $V\beta$ が実際に分類をおこないたい文書集合に適したものとなっていない場合に、 $V\gamma$ を考慮することで、分類対象文書にあわせた最適化をおこなうことができる。たとえば、図3（d）に示すようにカテゴリA～Cの特徴ベクトル空間 $V A \sim V C$ では、W6およびW7のベクトル長さはともに0であるが、分類不能カテゴリの特徴ベクトル空間 $V$ ではそれぞれ1.09、0.29であるので、W6やW7、すなわち訓練用文書集合202aには現れていなかった新たな語や新たな分野の語などを含む文書は、そのぶん分類不能カテゴリに分類される可能性が高くなる。

【0044】なお、上記では $V\alpha$ および $V\gamma$ の算出に、分類対象文書集合201aのすべての文書をもちいるようにしたが、その数が多いときは処理の負荷が大きいため、分類対象文書集合201aから無作為に抽出した一部の文書をもちいるようにしてもよい。

【0045】なお、この特徴ベクトル空間推定部202が、請求項にいう「第1の特徴ベクトル空間推定手段」「第2の特徴ベクトル空間推定手段」あるいは「特徴ベクトル空間推定手段」に相当し、そのおこなう処理の中に、請求項にいう「第1の特徴ベクトル空間推定工程」「第2の特徴ベクトル空間推定工程」あるいは「特徴ベクトル空間推定工程」が含まれる。

【0046】図2に戻り、残りの機能部について説明を続ける。203は、分類カテゴリ推定部であり、分類対象文書集合201aの各文書の特徴ベクトルを、特徴ベ

クトル空間推定部202で算出された各カテゴリの特徴ベクトル空間（これは分類用データベース202cに保存されている）と比較して、その類似度を順次算出する。なお、類似度の算出方法としてはTF・IDF、naive Bayes、最小2乗法、最大エントロピー法などの従来技術が多数存在し、ここではそのいずれを採用するのであってもよい。

【0047】そして、類似度のもっとも高いカテゴリを当該文書の分類先と推定し、文書とその分類先カテゴリとを対応づけて、図5に例示するような分類結果一覧表203aを作成する。なお、この分類カテゴリ推定部203が、請求項にいう「分類カテゴリ推定手段」に相当し、またそのおこなう処理が、請求項にいう「分類カテゴリ推定工程」に相当する。

【0048】204は、カテゴリ体系管理部であり、あらかじめカテゴリ体系表204aと、カテゴリ体系変更規則204bとを保持している。カテゴリ体系表204aでは、文書が分類される各カテゴリの名称と、その上位・下位関係とが定義されている。カテゴリの名称と関係とが分かるのであれば、カテゴリ体系表204aのデータ構造はどのようなものであってもよい。

【0049】また、カテゴリ体系変更規則204bとは、カテゴリ体系表204aに新たなカテゴリを追加したり、すでに登録されているカテゴリを分割・削除あるいは併合したりするのが望ましいかどうかを判定するための複数の規則（判定基準）である。

【0050】たとえば、カテゴリ追加に関しては「（a-1）分類不能と判定される頻度が閾値を上回った（具体的には「全文書中、分類不能カテゴリに属すると判定された文書の割合が20%を上回った」「分類不能カテゴリに属すると判定された文書が30以上連続して出現した」など）」という規則があり、この規則に該当する場合に、新たなカテゴリの追加が望ましいと判定する。

【0051】カテゴリ追加の規則としては、上記のほか「（a-2）分類不能カテゴリに属する文書の数閾値を上回った」、「（a-3）上記頻度および／または文書数をもちいた評価関数の結果が閾値を上回った（具体的には「分類不能カテゴリに属すると判定された文書の連続出現回数が、全文書数の15%にあたる数値を上回った」など）」、などがある。以下ではこれらの規則のうち、最初に挙げた「全文書中、分類不能カテゴリに属すると判定された文書の割合が20%を上回った」を、カテゴリ追加規則として採用する。

【0052】また、カテゴリ分割に関する規則としては、たとえば「（b-1）あるカテゴリが分類先であると判定される頻度が閾値を上回った（具体的には「全文書中、あるカテゴリに属すると判定された文書の割合が20%を上回った」「あるカテゴリに属すると判定された文書が30以上連続して出現した」など）」、「（b-2）あるカテゴリに属する文書の数閾値を上回っ



た」、「(b-3) 上記頻度および／または文書数もちいた評価関数の結果が閾値を上回った（具体的には「あるカテゴリに属すると判定された文書の連続出現回数が、全文書数の15%にあたる数値を上回った」など）」、などがある。以下ではこれらの規則のうち、最初に挙げた「全文書中、あるカテゴリに属すると判定された文書の割合が20%を上回った」を、カテゴリ分割規則として採用する。

【0053】また、カテゴリ削除または併合に関する規則としては、「(c-1) あるカテゴリが分類先であると判定される頻度が閾値を下回った（具体的には「全文書中、あるカテゴリに属すると判定された文書の割合が5%を下回った」「あるカテゴリ以外に分類される文書が50以上連続して出現した」など）」がある。ここでは最初に挙げた「全文書中、あるカテゴリに属すると判定された文書の割合が5%を下回った」を、カテゴリ削除・併合規則として採用する。

【0054】カテゴリ体系管理部204は、分類カテゴリ推定部203による文書の分類が終了すると、作成された分類結果一覧表203aを参照して、カテゴリの追加や分割、削除あるいは併合が望ましいかどうか、具体的には、カテゴリ体系変更規則204bの各規則に該当する事象が発生しているかどうかを判定する。

【0055】たとえば、分類結果一覧表203aが図5に示すようであり、カテゴリ体系変更規則204bは追加規則「全文書中、分類不能カテゴリに属すると判定された文書の割合が20%を上回った」の一つのみであったとすると、図5から分類不能カテゴリの文書数（ここでは2）が全文書数（ここでは5）の20%を上回ったことが分かるので、新たなカテゴリの追加が望ましいと判定する。カテゴリの分割や削除・併合についても、上記と同様にしてその可否を判定する。

【0056】なお、このカテゴリ体系管理部204が、請求項にいう「カテゴリ追加要否判定手段」「カテゴリ分割要否判定手段」あるいは「カテゴリ削除・併合要否判定手段」に相当し、またそのおこなう処理の中に、請求項にいう「カテゴリ追加要否判定工程」「カテゴリ分割要否判定工程」あるいは「カテゴリ削除・併合要否判定工程」が含まれる。

【0057】ユーザーインターフェース部205は、分類カテゴリ推定部203により作成された分類結果一覧表203a、およびカテゴリ体系管理部204のカテゴリ体系表204aを読み込んで、カテゴリのツリー構造と各カテゴリに分類される文書のタイトルとをグラフィカルに画面表示する。また、カテゴリ体系管理部204によってカテゴリの追加・分割・削除あるいは併合が望ましいと判定された場合に、カテゴリの追加などを実行するよう操作者に推奨するダイアログ、またその実行に際して必要なパラメータ（追加するカテゴリの名称など）を入力させるダイアログを表示する。これらの表示

例については後述する。

【0058】なお、このユーザーインターフェース部205が、請求項にいう「カテゴリ追加推奨手段」「カテゴリ分割推奨手段」あるいは「カテゴリ削除・併合推奨手段」に相当し、またそのおこなう処理の中に、請求項にいう「カテゴリ追加推奨工程」「カテゴリ分割推奨工程」あるいは「カテゴリ削除・併合推奨工程」が含まれる。

【0059】なお、文書記憶部201、特徴ベクトル空間推定部202、分類カテゴリ推定部203、カテゴリ体系管理部204およびユーザーインターフェース部205は、それぞれROM102、RAM103またはハードディスク105、フロッピーディスク107などの記録媒体に記録されたプログラムに記載された命令にしたがってCPU101などが命令処理を実行することにより、各部の機能を実現するものである。

【0060】つぎに、この発明の実施の形態による文書分類装置の文書分類処理の手順について説明する。図6は、この発明の実施の形態による文書分類装置の文書分類処理の手順を示すフローチャートである。分類対象文書集合201aに新たな文書が追加されたり、あるいは操作者から文書の分類が指示されたりした場合に、本フローチャートによる処理を開始する。

【0061】ステップS601において、特徴ベクトル空間推定部202は、訓練用文書集合202aの各文書についてその特徴ベクトルを作成する。そして、ステップS602で、文書-カテゴリ対応表202bの各カテゴリごとに文書の特徴ベクトルを平均して、各カテゴリの特徴ベクトル空間を推定（算出）する。

【0062】なお、これ以前にも本フローチャートによる処理をおこなったことがあれば（すなわち、2回目以降の分類処理の場合は）、以前の処理時にも訓練用文書集合202aにもとづく各カテゴリの特徴ベクトル空間の算出がおこなわれ、その結果が分類用データベース202cに保存されているので、ステップS601およびS602の処理は省略することができる。

【0063】さらに、ステップS603において、分類不能カテゴリの特徴ベクトル空間Vを推定する。これは上述のように、すべてのベクトルの長さが等しい特徴ベクトル空間V $\alpha$ （分類対象文書集合201aおよび訓練用文書集合202aの双方から算出）と、訓練用文書集合202aの特徴ベクトル空間V $\beta$ と、分類対象文書集合201aの特徴ベクトル空間V $\gamma$ との加重平均によって算出する。そして、ステップS604で、ステップS602およびステップS603で算出された各カテゴリの特徴ベクトル空間を、分類用データベース202cに保存する。

【0064】ステップS605で、分類カテゴリ推定部203は、分類対象文書集合201aの各文書についてその特徴ベクトルを作成する。なお、ステップS603

で $V\alpha$ や $V\gamma$ が分類対象文書の全部をもちいて算出された場合は、その際にすでにすべての文書の特徴ベクトルが作成されているので、このステップS605の処理は省略することができる。また、ステップS603で $V\alpha$ や $V\gamma$ が分類対象文書の一部をもちいて算出された場合には、そこに含まれなかった文書についてのみ特徴ベクトルを作成すればよい。

【0065】ステップS606で、分類カテゴリ推定部203は、ステップS605で作成された各文書の特徴ベクトルと、ステップS604で作成された分類用データベース202cの各カテゴリの特徴ベクトル空間とを順次比較して、もっとも類似度の高いカテゴリを当該文書の分類先と推定する。そして、ステップS607で、各文書とその分類先のカテゴリとを分類結果一覧表203aに保存する。

【0066】ステップS608で、ユーザーインターフェース部205は、ステップS607で作成された分類結果一覧表203aを読み込んで、図7に示すような分類結果一覧ウィンドウを表示する。そして、本フローチャートによる処理を終了する。

【0067】つぎに、この発明の実施の形態による文書分類装置のカテゴリ追加処理の手順について説明する。図8は、この発明の実施の形態による文書分類装置のカテゴリ追加処理の手順を示すフローチャートである。

【0068】図6に示す文書分類処理が終了した直後に、本フローチャートによる処理を開始する（あるいは、分類処理とは無関係にこの処理のみを定期的におこなうようにしてもよい）。なお、この開始時点で画面表示されている分類結果一覧ウィンドウは、図7に示すようなものであったとする。同図は分類不能カテゴリに大量の文書（たとえば全文書数の20%超）が分類されている状態を示している。

【0069】ステップS801で、カテゴリ体系管理部204は、図6のステップS607で作成された分類結果一覧表203aを参照して、カテゴリ体系変更規則204bのカテゴリ追加規則に該当する事象が発生しているかどうかを判定する。そして、カテゴリ追加規則に該当する事象が発生しているときは（ステップS801肯定）ステップS802に移行し、発生していないときは（ステップS801否定）そのまま本フローチャートによる処理を終了する。

【0070】ステップS802で、ユーザーインターフェース部205は、図9に示すような変更推奨ダイアログを表示する。ここでは新たなカテゴリの追加を推奨している。そして、ステップS803でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが変更開始ボタン901であったときは（ステップS803肯定）ステップS804に移行する。また、マウスクリックされたのがキャンセルボタン902であったときは（ステップS803否定）、そのまま本フロー

チャートによる処理を終了する。

【0071】ステップS804で、ユーザーインターフェース部205は、ステップS802で表示された変更推奨ダイアログを消去するとともに、図10に示すようなカテゴリ追加ダイアログを表示する。これは追加する新たなカテゴリの名称を操作者に入力させるためのダイアログである。そして、ステップS805でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが「進む」ボタン1001であったときは（ステップS805肯定）ステップS807に移行し、「進む」ボタン1001でなかったときは（ステップS805否定）、ステップS806に移行して、クリックされたその他のボタンに応じた処理をおこなう。

【0072】ステップS807で、ステップS804で表示されたカテゴリ追加ダイアログを消去するとともに、図11に示すような文書割付ダイアログを表示する。これは、追加する新たなカテゴリ（ここでは情報家電カテゴリ）に分類する文書を操作者に指定させるためのダイアログであり、分類不能カテゴリに分類されているすべての文書が一覧表示される。そして、ステップS808でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが「完了」ボタン1101であったときは（ステップS808肯定）ステップS810に移行し、「完了」ボタン1101でなかったときは（ステップS808否定）、ステップS809に移行して、クリックされたその他のボタンに応じた処理をおこなう。

【0073】ステップS810で、ユーザーインターフェース部205は、ステップS805で「進む」ボタン1001がクリックされた時点でカテゴリ追加ダイアログに入力されていたカテゴリ名称（ここでは「情報家電」）を、カテゴリ体系管理部204に対して通知する。これを受けたカテゴリ体系管理部204は、カテゴリ体系表204aに新たなカテゴリ「情報家電」を追記する。

【0074】また、ステップS811で、ユーザーインターフェース部205は、ステップS808で文書割付ダイアログの「完了」ボタン1101がクリックされた時点で選択されていた文書（ここでは分類不能カテゴリのすべての文書が選択されていたものとする）を、特徴ベクトル空間推定部202に対して通知する。これを受けた特徴ベクトル空間推定部202は、通知された文書について図6のステップS605で作成された特徴ベクトルを取得して、それらの平均を取ることで、追加するカテゴリの特徴ベクトル空間を算出する。そして、これをカテゴリ名称と対応づけて、分類用データベース202cに登録する。

【0075】さらに、ステップS812で、ステップS807で表示された文書割付ダイアログを消去した後、図6のステップS606に移行して、更新された分類用

データベース202cをもとに分類対象文書集合201aの分類を再度実行する。図12は、新たなカテゴリ「情報家電」が追加された直後の分類結果一覧ウィンドウの一例を示す説明図である。図7で分類不能カテゴリに分類されていた文書が、図12では情報家電カテゴリに分類されている。

【0076】つぎに、この発明の実施の形態による文書分類装置のカテゴリ分割処理の手順について説明する。図13は、この発明の実施の形態による文書分類装置のカテゴリ分割処理の手順を示すフローチャートである。図8に示すカテゴリ追加処理が終了した直後に、本フローチャートによる処理を開始する。なお、この開始時点で画面表示されている分類結果一覧ウィンドウは、図14に示すようなものであったとする。同図は携帯端末カテゴリに大量の文書（たとえば全文書数の20%超）が分類されている状態を示している。

【0077】ステップS1301で、カテゴリ体系管理部204は、図8のステップS811で更新された分類結果一覧表203aを参照して、カテゴリ体系変更規則204bのカテゴリ分割規則に該当する事象が発生しているかどうかを判定する。そして、カテゴリ分割規則に該当する事象が発生しているときは（ステップS1301肯定）ステップS1302に移行し、発生していないときは（ステップS1301否定）そのまま本フローチャートによる処理を終了する。

【0078】ステップS1302で、ユーザーインターフェース部205は、図15に示すような変更推奨ダイアログを表示する。ここでは携帯端末カテゴリの分割を推奨している。そして、ステップS1303でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが変更開始ボタン1501であったときは（ステップS1303肯定）ステップS1304に移行する。また、マウスクリックされたのがキャンセルボタン1502であったときは（ステップS1303否定）、そのまま本フローチャートによる処理を終了する。

【0079】ステップS1304で、ユーザーインターフェース部205は、ステップS1302で表示された変更推奨ダイアログを消去するとともに、図16に示すようなカテゴリ分割ダイアログを表示する。これは分割後の各カテゴリの名称を操作者に入力させるためのダイアログである。そして、ステップS1305でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが「進む」ボタン1601であったときは（ステップS1305肯定）ステップS1307に移行し、「進む」ボタン1601でなかったときは（ステップS1305否定）、ステップS1306に移行して、クリックされたその他のボタンに応じた処理をおこなう。

【0080】ステップS1307で、ステップS130

4で表示されたカテゴリ分割ダイアログを消去するとともに、図17に示すような文書割付ダイアログを表示する。これは、分割後のカテゴリの一方（ここでは携帯電話カテゴリ）に分類する文書を操作者に指定させるためのダイアログであり、分割前のカテゴリ（ここでは携帯端末カテゴリ）に分類されているすべての文書が一覧表示される。

【0081】そして、ステップS1308でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが「進む」ボタン1701であったときは（ステップS1308肯定）ステップS1310に移行し、「進む」ボタン1701でなかったときは（ステップS1308否定）、ステップS1309に移行して、クリックされたその他のボタンに応じた処理をおこなう。

【0082】ステップS1310で、ステップS1307で表示された文書割付ダイアログを消去するとともに、同様の文書割付ダイアログを分割後のカテゴリのもう一方（ここでは携帯PCカテゴリ）について表示する。ただし、このダイアログには分割前のカテゴリの全文書から、ステップS1308で分割後の他のカテゴリに分類された文書を除いた文書の一覧が表示される。そして、ステップS1311でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが「完了」ボタン1702であったときは（ステップS1311肯定）ステップS1313に移行し、「完了」ボタン1702でなかったときは（ステップS1311否定）、ステップS1312に移行して、クリックされたその他のボタンに応じた処理をおこなう。

【0083】ステップS1313で、ユーザーインターフェース部205は、分割前のカテゴリ名称（ここでは「携帯端末」）および分割後のカテゴリ名称（ステップS1305で「進む」ボタン1601がクリックされた時点でカテゴリ分割ダイアログに入力されていたカテゴリ名称。ここでは「携帯電話」「携帯PC」）を、カテゴリ体系管理部204に対して通知する。これを受けたカテゴリ体系管理部204は、カテゴリ体系表204aからカテゴリ「携帯端末」を削除するとともに、新たなカテゴリ「携帯電話」および「携帯PC」を追記する。

【0084】また、ステップS1314で、ユーザーインターフェース部205は、分割前後のカテゴリ名称およびステップS1308およびステップS1311で分割後の各カテゴリに割り付けられた文書を、特徴ベクトル空間推定部202に対して通知する。これを受けた特徴ベクトル空間推定部202は、分割前のカテゴリの特徴ベクトル空間を分類用データベース202cから削除するとともに、分割後の各カテゴリの特徴ベクトル空間を算出して、カテゴリ名称と対応づけて上記データベースに登録する。

【0085】さらに、ステップS1315で、ステップ

S1310で表示された文書割付ダイアログを消去した後、図6のステップS606に移行して、更新された分類用データベース202cをもとに分類対象文書集合201aの分類を再度実行する。図18は、携帯端末カテゴリが携帯電話カテゴリおよび携帯PCカテゴリに分割された直後の分類結果一覧ウィンドウの一例を示す説明図である。

【0086】図14で携帯端末カテゴリに分類されていた文書の一部が、図18では携帯PCカテゴリに分類されている。また、カテゴリ分割後に再度分類対象文書集合201aの全文書について分類をおこなった結果、分割前の携帯端末カテゴリには含まれていなかった文書

(たとえば「携帯PCの音声入力規格をI社が提案。J社なども賛同」)が、分割後の携帯PCカテゴリに含まれるようになっていく。なお、図示は省略するが、携帯電話カテゴリも上記と同様である。なお、上記ではカテゴリを2つに分割したが、この個数はいくつであってもよい。

【0087】つぎに、この発明の実施の形態による文書分類装置のカテゴリ削除・併合処理の手順について説明する。図19は、この発明の実施の形態による文書分類装置のカテゴリ削除・併合処理の手順を示すフローチャートである。図13に示すカテゴリ分割処理が終了した直後に、本フローチャートによる処理を開始する。なお、この開始時点で画面表示されている分類結果一覧ウィンドウは、図20に示すようなものであったとする。同図はiNet専用端末カテゴリに分類された文書が少ない(たとえば全文書数の5%未満)状態を示している。

【0088】ステップS1901で、カテゴリ体系管理部204は、図13のステップS1314で更新された分類結果一覧表203aを参照して、カテゴリ体系変更規則204bのカテゴリ削除・併合規則に該当する事象が発生しているかどうかを判定する。そして、カテゴリ削除・併合規則に該当する事象が発生しているときは

(ステップS1901肯定) ステップS1902に移行し、発生していないときは(ステップS1901否定)そのまま本フローチャートによる処理を終了する。

【0089】ステップS1902で、ユーザーインターフェース部205は、図21に示すような変更推奨ダイアログを表示する。ここではiNet専用端末カテゴリの削除または併合を推奨している。そして、ステップS1903でいずれかのボタンがマウスクリックされるのを待ち、マウスクリックされたのが変更開始ボタン2101であったときは(ステップS1903肯定) ステップS1904に移行する。また、マウスクリックされたのがキャンセルボタン2102であったときは(ステップS1903否定)、そのまま本フローチャートによる処理を終了する。

【0090】ステップS1904で、ユーザーインター

フェース部205は、ステップS1902で表示された変更推奨ダイアログを消去するとともに、図22に示すようなカテゴリ削除・併合ダイアログを表示する。これは文書数の少ないカテゴリを削除するのか、あるいは併合するのかを操作者に指定させるためのダイアログである。なお、削除が指定されている間は「進む」ボタン2201を、併合が指定されている間は「完了」ボタン2202を、それぞれグレースアウトして押下できないようにする。

【0091】そして、ステップS1905でいずれかのボタンが押下されるのを待ち、押下されたのが「完了」ボタン2202であったときは(ステップS1905肯定)、カテゴリの削除が指定されたので、削除するカテゴリの名称(ここでは「iNet専用端末」)をステップS1906でカテゴリ体系管理部204に通知して、カテゴリ体系表204aから削除させる。また、ステップS1907で特徴ベクトル空間推定部202にも通知して、文書-カテゴリ対応表202bおよび分類用データベース202cから削除させる。そして、ステップS1908で、ステップS1904で表示されたカテゴリ削除・併合ダイアログを消去した後、図6のステップS606に移行する。

【0092】また、ステップS1905で押下されたのが「完了」ボタン2202でなかったときは(ステップS1905否定)、ステップS1909に移行して、それが「進む」ボタン2201であったかどうかを判定する。そして、「進む」ボタン2201でもなかったときは(ステップS1909否定)、ステップS1910で押下されたその他のボタンに応じた処理をおこなうが、「進む」ボタン2201であったときは(ステップS1909肯定)、ステップS1911に移行する。

【0093】この場合はカテゴリの併合が指定されたので、ステップS1911で、ステップS1904で表示されたカテゴリ削除・併合ダイアログを消去するとともに、図23に示すような併合先選択ダイアログを表示する。これは、併合により削除されるカテゴリの文書を他のどのカテゴリに分類するかを操作者に指定させるためのダイアログであり、併合されるカテゴリ以外のすべてのカテゴリが一覧表示される。

【0094】そして、ステップS1912でいずれかのボタンが押下されるのを待ち、押下されたのが「進む」ボタン2301でなかったときは(ステップS1912否定)、ステップS1913で押下されたその他のボタンに応じた処理をおこなうが、「進む」ボタン2301であったときは(ステップS1912肯定)、ステップS1914に移行する。

【0095】ステップS1914で、ステップS1911で表示した併合先選択ダイアログを消去するとともに、図24に示すような文書割付ダイアログを表示する。これは、併合により削除されるカテゴリの文書のう

ちいずれの文書を、併合先のカテゴリ（ステップS1912で「進む」ボタン2301が押下された時点で選択されていたカテゴリ。ここでは情報家電カテゴリであったものとする）に分類するかを操作者に指定させるためのダイアログであり、併合されるカテゴリのすべての文書の一覧が表示される。

【0096】そして、ステップS1915でいずれかのボタンが押下されるのを待ち、押下されたのが「完了」ボタン2401でなかったときは（ステップ1915否定）、ステップS1916で押下されたその他のボタン

に応じた処理をおこなうが、「完了」ボタン2401であったときは（ステップS1915肯定）、ステップS1917に移行する。

【0097】ステップS1917で、ユーザーインターフェース部205は、併合されるカテゴリの名称（ここでは「iNet専用端末」）をカテゴリ体系管理部204に通知して、カテゴリ体系表204aから削除させる。また、ステップS1918で、併合元および併合先のカテゴリ名称、および併合先に割り付けられる文書

（ステップS1915で「完了」ボタン2401が押下された時点で選択されていた文書）を、特徴ベクトル空間推定部202に対して通知する。これを受けた特徴ベクトル空間推定部202は、文書-カテゴリ対応表202bの併合元カテゴリを併合先カテゴリに書き換え、併合先のカテゴリの特徴ベクトル空間を再計算して分類用データベース202cに登録するとともに、併合元のカテゴリの特徴ベクトル空間を上記データベースから削除する。

【0098】そして、ステップS1919で、ステップS1914で表示された文書割付ダイアログを消去した

後、図6のステップS606に移行して、更新された分類用データベース202cをもとに分類対象文書集合201aの分類を再度実行する。図25は、iNet専用端末カテゴリが削除、あるいは情報家電カテゴリに併合された直後の分類結果一覧ウィンドウの一例を示す説明図である。

【0099】以上説明したようにこの実施の形態によれば、文書の追加や文書の分類指示によって文書の分類を実行するたびに、特定のカテゴリの文書（分類不能カテゴリを含む）の文書が増えすぎているかどうかや、逆に特定のカテゴリの文書が減りすぎているかなどが自動的にチェックされ、そのチェック結果にもとづいてカテゴリの追加や分割、削除、併合のいずれかが推奨されるので、この推奨にしたがって必要事項などを入力してゆくだけで、分類対象文書の特性に見合った最適な分類体系を常に維持することができる（分類体系のメンテナンスの容易化）。したがって、分類対象文書の分野傾向が変わるなどの質的変化があった場合にも、常に適切な分類結果を得ることが可能である。

【0100】また、分類不能カテゴリの特徴ベクトル空

間Vを、すべてのベクトルの長さが等しい特徴ベクトル空間 $V\alpha$ 、訓練用文書集合202aの特徴ベクトル空間 $V\beta$ および分類対象文書集合201aの特徴ベクトル空間 $V\gamma$ の加重平均によって推定するので、既存のカテゴリとの類似度の低い文書が出現した場合にも、まったく新しい分野のカテゴリを生成してその中に分類することが可能である。

【0101】なお、上記実施の形態で説明した文書分類方法（カテゴリ追加方法、カテゴリ分割方法およびカテゴリ削除・併合方法を含む）は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータで実行することにより実現される。このプログラムは、ハードディスク、フロッピーディスク、CD-ROM、MO、DVDなどのコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されて実行される。また、このプログラムは、上記記録媒体を介して、インターネットなどのネットワークを介して配布することができる。

【0102】

【発明の効果】以上説明したように請求項1に記載の発明にかかる文書分類装置は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置において、第1の特徴ベクトル空間推定手段が、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定し、第2の特徴ベクトル空間推定手段が、すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定し、分類カテゴリ推定手段が、前記第1の特徴ベクトル空間推定手段により推定された各カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定手段により推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定し、カテゴリ追加要否判定手段が、前記分類カテゴリ推定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定し、カテゴリ追加推奨手段が、前記カテゴリ追加要否判定手段により分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するので、分類不能カテゴリに分類される文書が多くなると、操作者に対して新たなカテゴリの追加が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類装置が得られるという効果を奏する。



【0103】また、請求項2に記載の発明にかかる文書分類装置は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類装置において、特徴ベクトル空間推定手段が、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定し、分類カテゴリ推定手段が、前記特徴ベクトル空間推定手段により推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定し、カテゴリ削除・併合要否判定手段が、前記分類カテゴリ推定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定し、カテゴリ削除・併合推奨手段が、前記カテゴリ削除・併合要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するので、あるカテゴリに分類される文書が少なくなると、操作者に対して当該カテゴリの削除または併合が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類装置が得られるという効果を奏する。

【0104】また、請求項3に記載の発明にかかる文書分類装置は、前記請求項1または2に記載の発明において、さらに、前記分類カテゴリ推定手段により分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定手段と、前記カテゴリ分割要否判定手段により分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨手段と、を備えたので、あるカテゴリに分類される文書が多くなると、操作者に対して当該カテゴリの分割が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類装置が得られるという効果を奏する。

【0105】また、請求項4に記載の発明にかかる文書分類方法は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、第1の特徴ベクトル空間推定工程で、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定し、第2の特徴ベクトル空間推定工程で、すべてのベクトルの長さが等しい特徴ベクトル空間、訓練用文書集合の特徴ベクトル空間および分類対象文書集合の特徴ベクトル空間の重み付き平均から分類不能カテゴリの特徴ベクトル空間を推定し、分類カテゴリ推定工程で、前記第1の特徴ベクトル空間推

定工程で推定された各カテゴリの特徴ベクトル空間および前記第2の特徴ベクトル空間推定工程で推定された分類不能カテゴリの特徴ベクトル空間と、分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリまたは前記分類不能カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定し、カテゴリ追加要否判定工程で、前記分類カテゴリ推定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったかどうかを判定し、カテゴリ追加推奨工程で、前記カテゴリ追加要否判定工程で分類不能カテゴリが分類先のカテゴリと推定される頻度が一定の閾値を上回ったと判定された場合に、新たなカテゴリの追加を操作者に対して推奨するので、分類不能カテゴリに分類される文書が多くなると、操作者に対して新たなカテゴリの追加が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類方法が得られるという効果を奏する。

【0106】また、請求項5に記載の発明にかかる文書分類方法は、電子化された文書をあらかじめ設定されたカテゴリのうちいずれか一つに分類する文書分類方法において、特徴ベクトル空間推定工程で、訓練用文書集合を構成する各文書の特徴ベクトルから前記各カテゴリの特徴ベクトル空間を推定し、分類カテゴリ推定工程で、前記特徴ベクトル空間推定工程で推定された各カテゴリの特徴ベクトル空間と分類対象文書集合を構成する各文書の特徴ベクトルとを比較することにより、前記各カテゴリのうちいずれか一つを前記各文書の分類先のカテゴリと推定し、カテゴリ削除・併合要否判定工程で、前記分類カテゴリ推定工程で分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあるかどうかを判定し、カテゴリ削除・併合推奨工程で、前記カテゴリ削除・併合要否判定工程で分類先のカテゴリと推定される頻度が一定の閾値を下回ったカテゴリがあると判定された場合に、当該カテゴリの削除または併合を操作者に対して推奨するので、あるカテゴリに分類される文書が少なくなると、操作者に対して当該カテゴリの削除または併合が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類方法が得られるという効果を奏する。

【0107】また、請求項6に記載の発明にかかる文書分類方法は、前記請求項4または5に記載の発明において、さらに、前記分類カテゴリ推定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあるかどうかを判定するカテゴリ分割要否判定工程と、前記カテゴリ分割要否判定工程で分類先のカテゴリと推定される頻度が一定の閾値を上回ったカテゴリがあ



ると判定された場合に、当該カテゴリの分割を操作者に対して推奨するカテゴリ分割推奨工程と、を含んだので、あるカテゴリに分類される文書が多くなると、操作者に対して当該カテゴリの分割が推奨され、これによって、分類対象文書群の特性に見合った分類体系を常に維持することができ、したがってその変化にともなう分類精度の低下を防止することが可能な文書分類方法が得られるという効果を奏する。

【0108】また、請求項7の発明にかかる記録媒体は、請求項4～6のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことで、当該プログラムをコンピュータで読み取ることが可能となり、これによって、請求項4～6のいずれか一つに記載された方法をコンピュータによって実施することが可能な記録媒体が得られるという効果を奏する。

#### 【図面の簡単な説明】

【図1】この発明の実施の形態による文書分類装置のハードウェア構成を示す説明図である。

【図2】この発明の実施の形態による文書分類装置の構成を機能的に示す説明図である。

【図3】この発明の実施の形態による文書分類装置の、特徴ベクトル空間推定部202による分類用データベース202cの作成方法を具体的に説明するための説明図である。

【図4】この発明の実施の形態による文書分類装置の、分類対象文書集合201aおよびその各文書について作成される特徴ベクトルの一例を示す説明図である。

【図5】この発明の実施の形態による文書分類装置の、分類カテゴリ推定部203により作成される分類結果一覧表203aの一例を示す説明図である。

【図6】この発明の実施の形態による文書分類装置の、文書分類処理の手順を示すフローチャートである。

【図7】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により作成される分類結果一覧ウィンドウの一例を示す説明図である。

【図8】この発明の実施の形態による文書分類装置の、カテゴリ追加処理の手順を示すフローチャートである。

【図9】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される変更推奨ダイアログの一例を示す説明図である。

【図10】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示されるカテゴリ追加ダイアログの一例を示す説明図である。

【図11】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される文書割付ダイアログの一例を示す説明図である。

【図12】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される分類結果一覧ウィンドウの他の一例を示す説明図である。

【図13】この発明の実施の形態による文書分類装置の、カテゴリ分割処理の手順を示すフローチャートである。

【図14】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される分類結果一覧ウィンドウの他の一例を示す説明図である。

【図15】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される変更推奨ダイアログの他の一例を示す説明図である。

【図16】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示されるカテゴリ分割ダイアログの一例を示す説明図である。

【図17】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される文書割付ダイアログの他の一例を示す説明図である。

【図18】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される分類結果一覧ウィンドウの他の一例を示す説明図である。

【図19】この発明の実施の形態による文書分類装置の、カテゴリ削除・併合処理の手順を示すフローチャートである。

【図20】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される分類結果一覧ウィンドウの他の一例を示す説明図である。

【図21】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される変更推奨ダイアログの他の一例を示す説明図である。

【図22】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示されるカテゴリ削除・併合ダイアログの一例を示す説明図である。

【図23】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される併合先選択ダイアログの一例を示す説明図である。

【図24】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される文書割付ダイアログの他の一例を示す説明図である。

【図25】この発明の実施の形態による文書分類装置の、ユーザーインターフェース部205により表示される分類結果一覧ウィンドウの他の一例を示す説明図である。

#### 【符号の説明】

100 バス

101 CPU

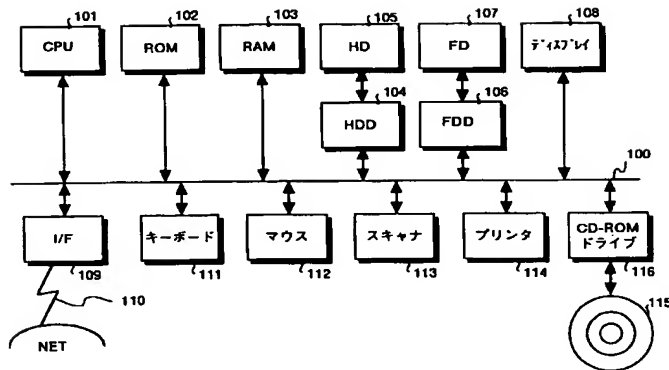
102 ROM

103 RAM

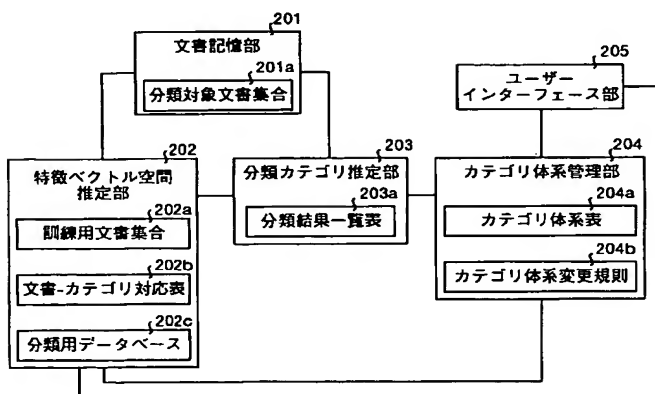
104 HDD

105 HD  
 106 FDD  
 107 FD  
 108 ディスプレイ  
 109 I/F  
 110 通信回線  
 111 キーボード  
 112 マウス  
 113 スキャナ  
 114 プリンタ  
 115 CD-ROM  
 116 CD-ROMドライブ

【図1】



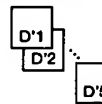
【図2】



201 文書記憶部  
 201a 分類対象文書集合  
 202 特徴ベクトル空間推定部  
 202a 訓練用文書集合  
 202b 文書-カテゴリ対応表  
 202c 分類用データベース  
 203 分類カテゴリ推定部  
 203a 分類結果一覧表  
 204 カテゴリ体系管理部  
 204a カテゴリ体系表  
 204b カテゴリ体系変更規則  
 205 ユーザーインターフェース部

【図4】

(a) 分類対象文書集合



(b) 分類対象文書の特徴ベクトル

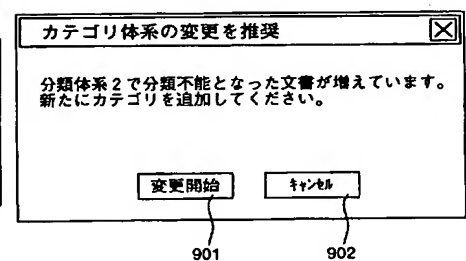
文書	単語							合計
	W1	W2	W3	W4	W5	W6	W7	
D'1	0	0	3	1	3	0	0	7
D'2	1	0	0	3	4	0	0	8
D'3	1	3	2	0	2	3	0	11
D'4	3	2	1	6	5	0	0	17
D'5	2	1	2	0	4	7	0	16
合計	7	6	8	10	18	10	0	59

【図5】

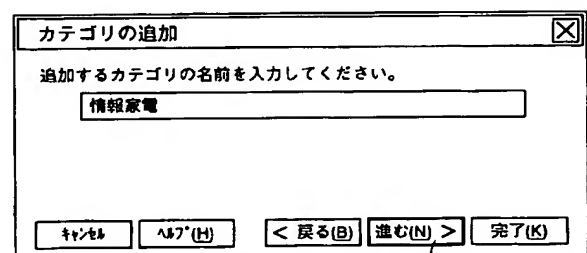
分類結果一覧表

文書	カテゴリ
D'1	C
D'2	B
D'3	分類不能
D'4	B
D'5	分類不能

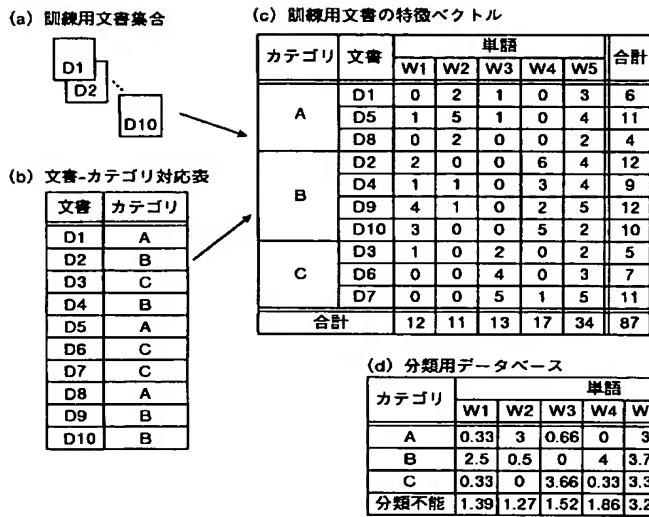
【図9】



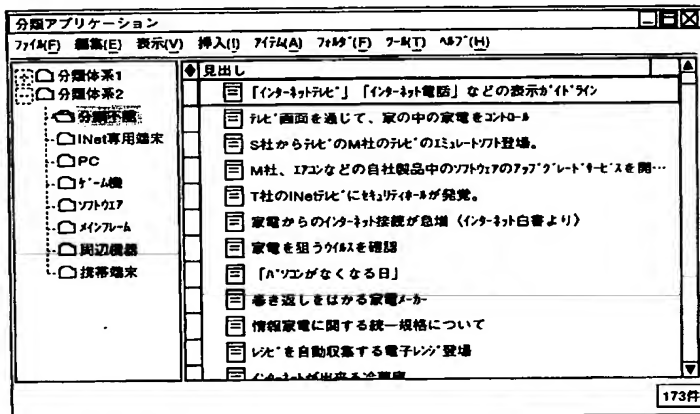
【図10】



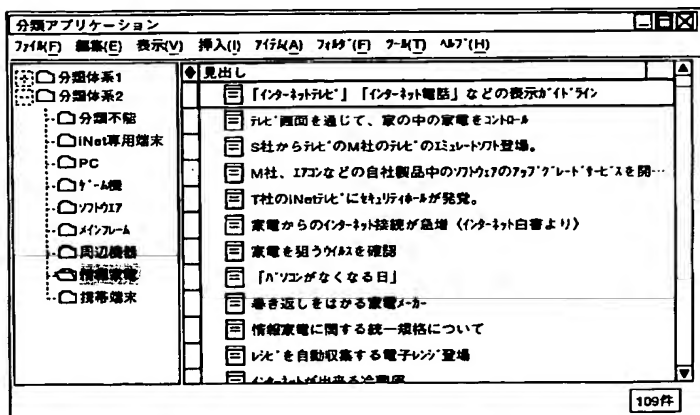
【図3】



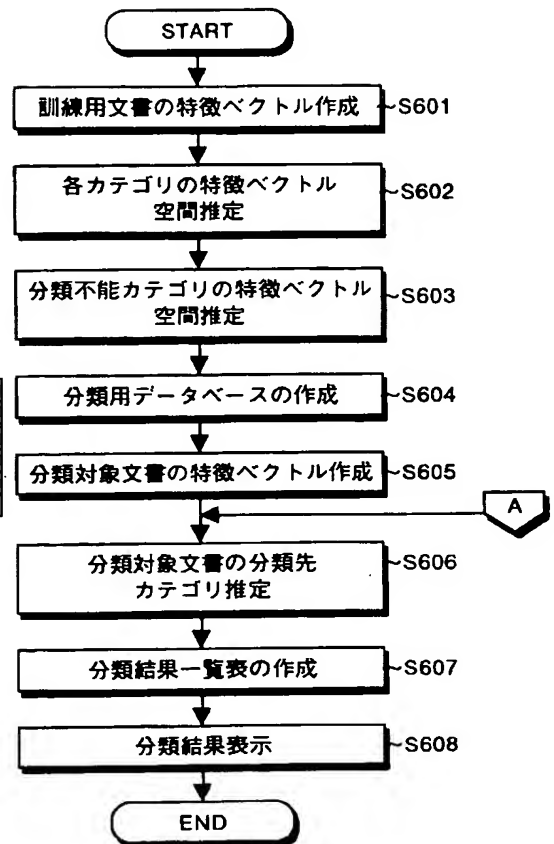
【図7】



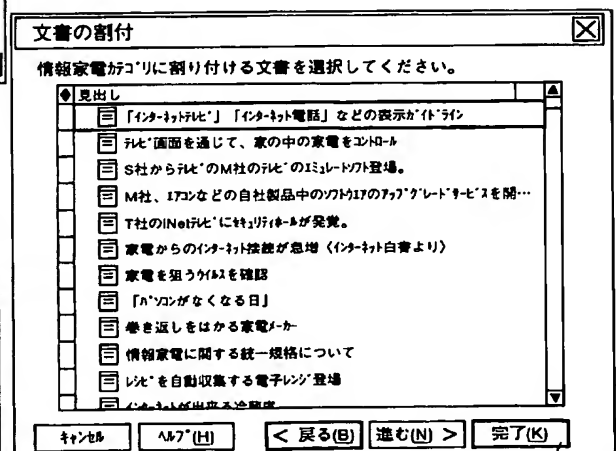
【図12】



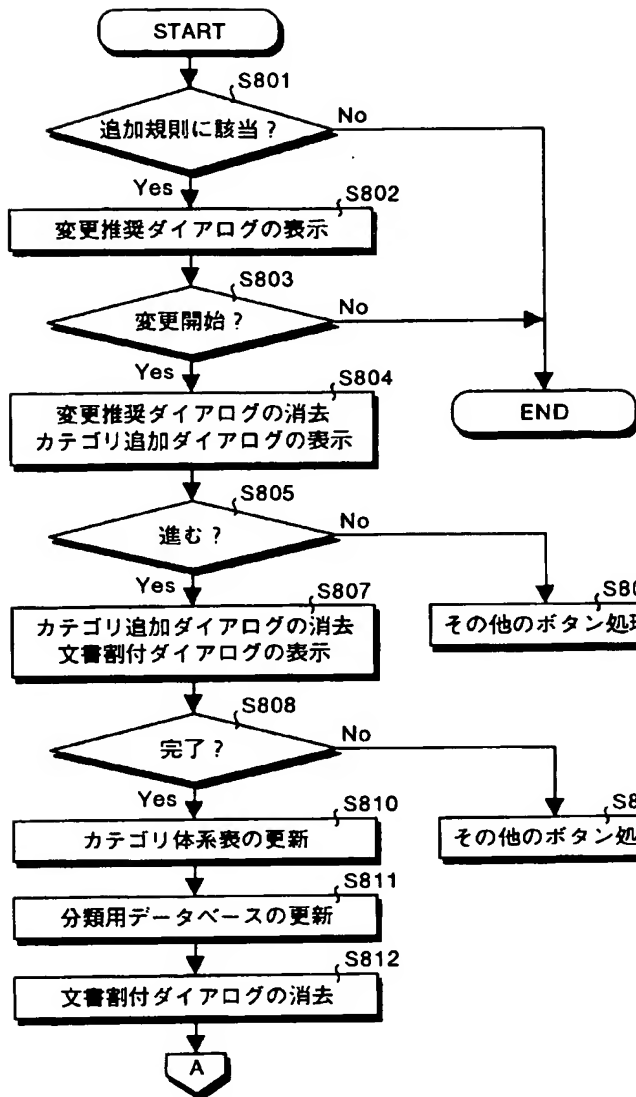
【図6】



【図11】



【図8】



【図16】

カテゴリの分割

分割の結果として追加するカテゴリの名前を入力してください。

携帯電話

追加(A) 除外(B)

携帯PC

キャンセル 実行(H) < 戻る(B) 進む(N) > 完了(K)

1601

【図15】

カテゴリ体系の変更を推奨

分類体系2で携帯端末カテゴリに分類された文書が多くなっています。さらに詳細なカテゴリに分割してください。

変更開始 1501 キャンセル 1502

【図17】

文書の割付

携帯電話カテゴリに割り付ける文書を選択してください。

見出し

- ☒ G社携帯電話の取り扱い。被害広がる。
- ☐ どちらを選ぶ？高性能携帯電話とインターネット電話付き携帯PC
- ☐ M社とN社、PC小型化技術に関して技術提携。顧客適合との声も。
- ☐ 携帯電話そっくりのPC発売。インターネット電話もできる。
- ☐ GPS内蔵パルムPCが専用機とともに登場
- ☐ 携帯型PDAから携帯型PCへ。
- ☐ 高性能PDA登場。
- ☐ 第4世代携帯電話に向けて各社の利害が対立
- ☐ 大学での講義中に携帯電話で通学処分。
- ☐ ネットワークのネットワークが携帯電話に人気
- ☐ PHS、携帯電話との差別化を図るも、なお、困難は続く
- ☐ 携帯電話の文字数制限が大幅緩和

キャンセル 実行(H) < 戻る(B) 進む(N) > 完了(K)

1701 1702

【図18】

分類アプリケーション

ファイル(F) 編集(E) 表示(V) 挿入(I) 実行(A) ヘルプ(H)

分類体系1

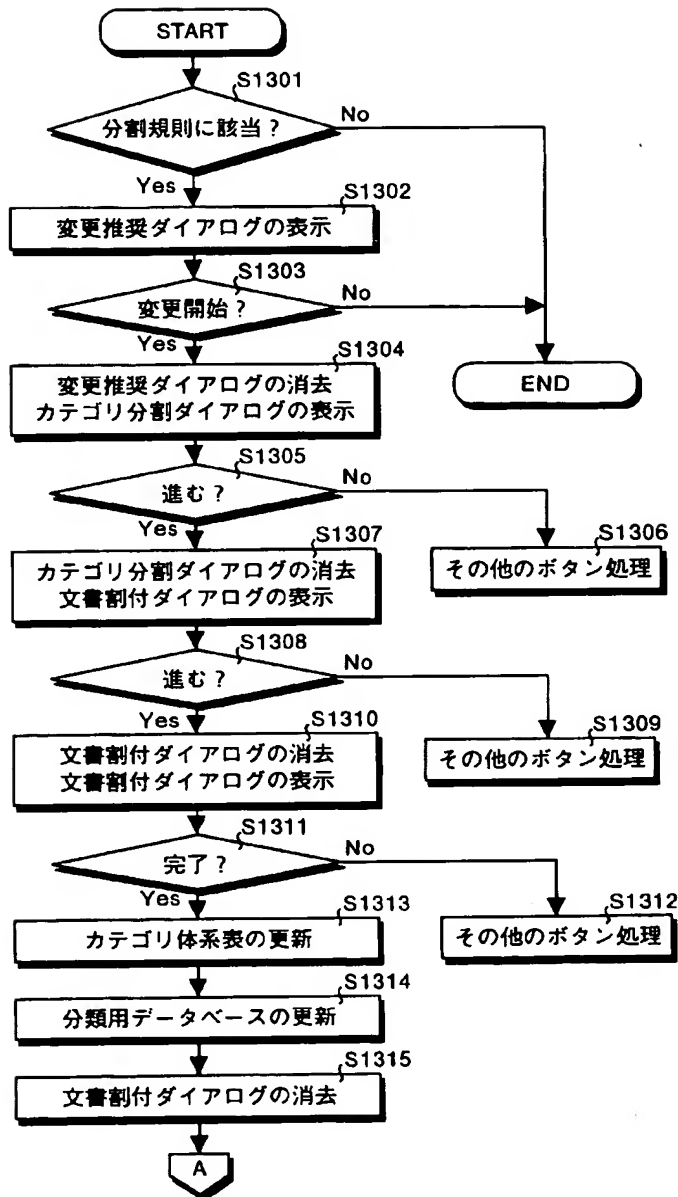
- ☒ 分類体系2
- ☐ 分類不能
- ☐ INet専用端末
- ☐ PC
- ☐ PDA
- ☐ パソコン
- ☐ メインフレーム
- ☐ 周辺機器
- ☐ 情報家電
- ☐ 携帯PC
- ☐ 携帯電話

見出し

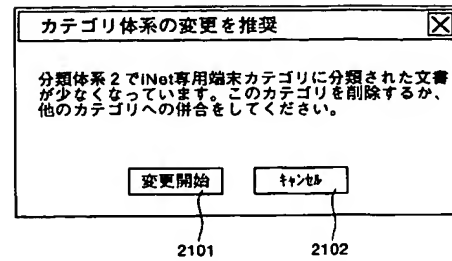
- ☒ 携帯PCの音声入力規格をI社が提案。J社なども賛同
- ☐ F社、携帯PC用ネットワーク開発。日本語の特性を活かして快適入力。
- ☐ PCの小型化。ハードウェアの時代か?
- ☐ 携帯PC、電子ブックや電子手帳との融合が進む。
- ☐ ネットワークにPCを
- ☐ 携帯PC、小型化と使いやすさの両立が最大の技術的課題
- ☐ どちらを選ぶ？高性能携帯電話とインターネット電話付き携帯PC
- ☐ M社とN社、PC小型化技術に関して技術提携。顧客適合との声も。
- ☐ 携帯電話そっくりのPC発売。インターネット電話もできる。
- ☐ GPS内蔵パルムPCが専用機とともに登場
- ☐ 携帯型PDAから携帯型PCへ。
- ☐ 高性能PDA登場。

73件

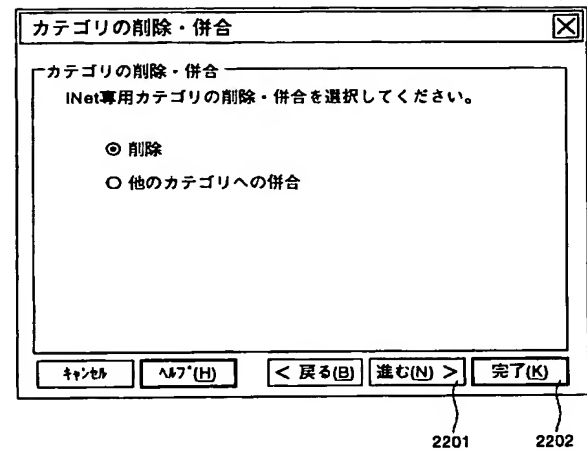
【図13】



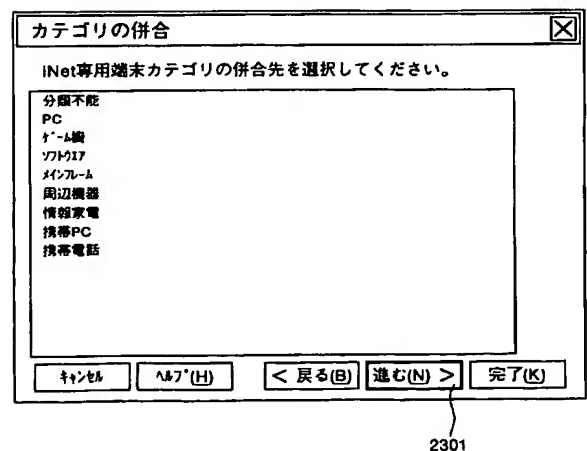
【図21】



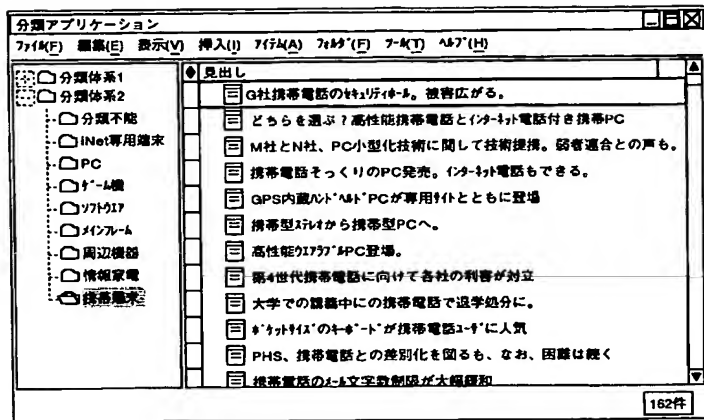
【図22】



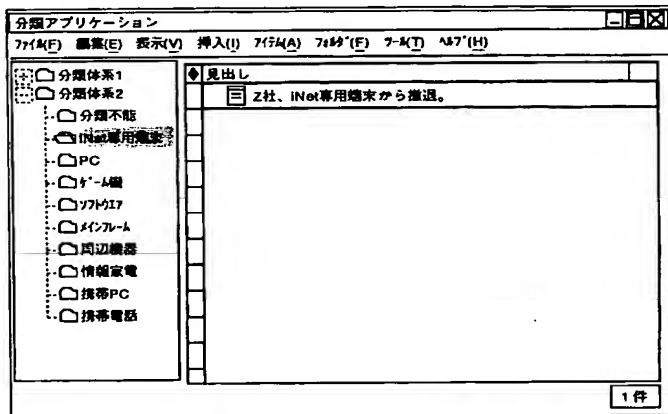
【図23】



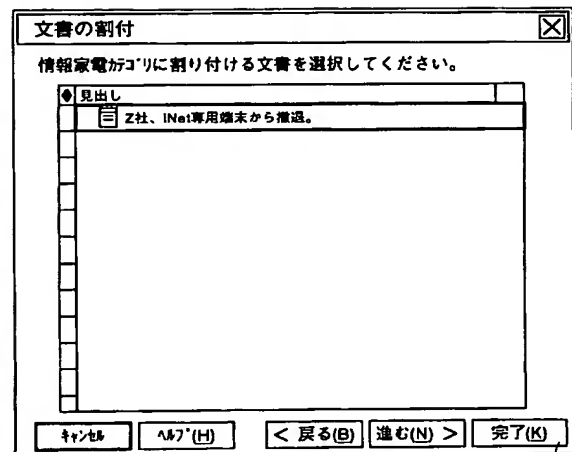
【図14】



【図20】

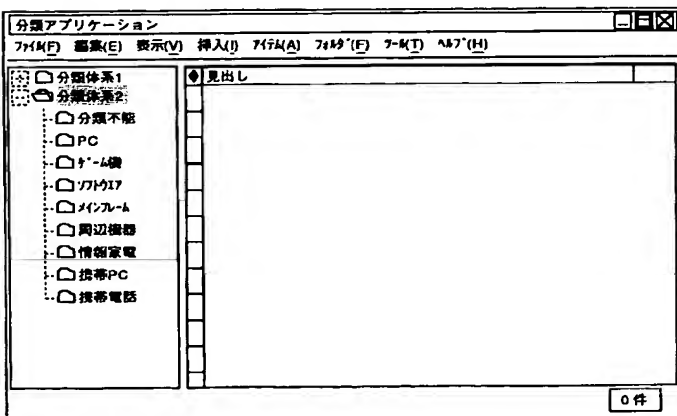


【図24】



2401

【図25】





【図19】

